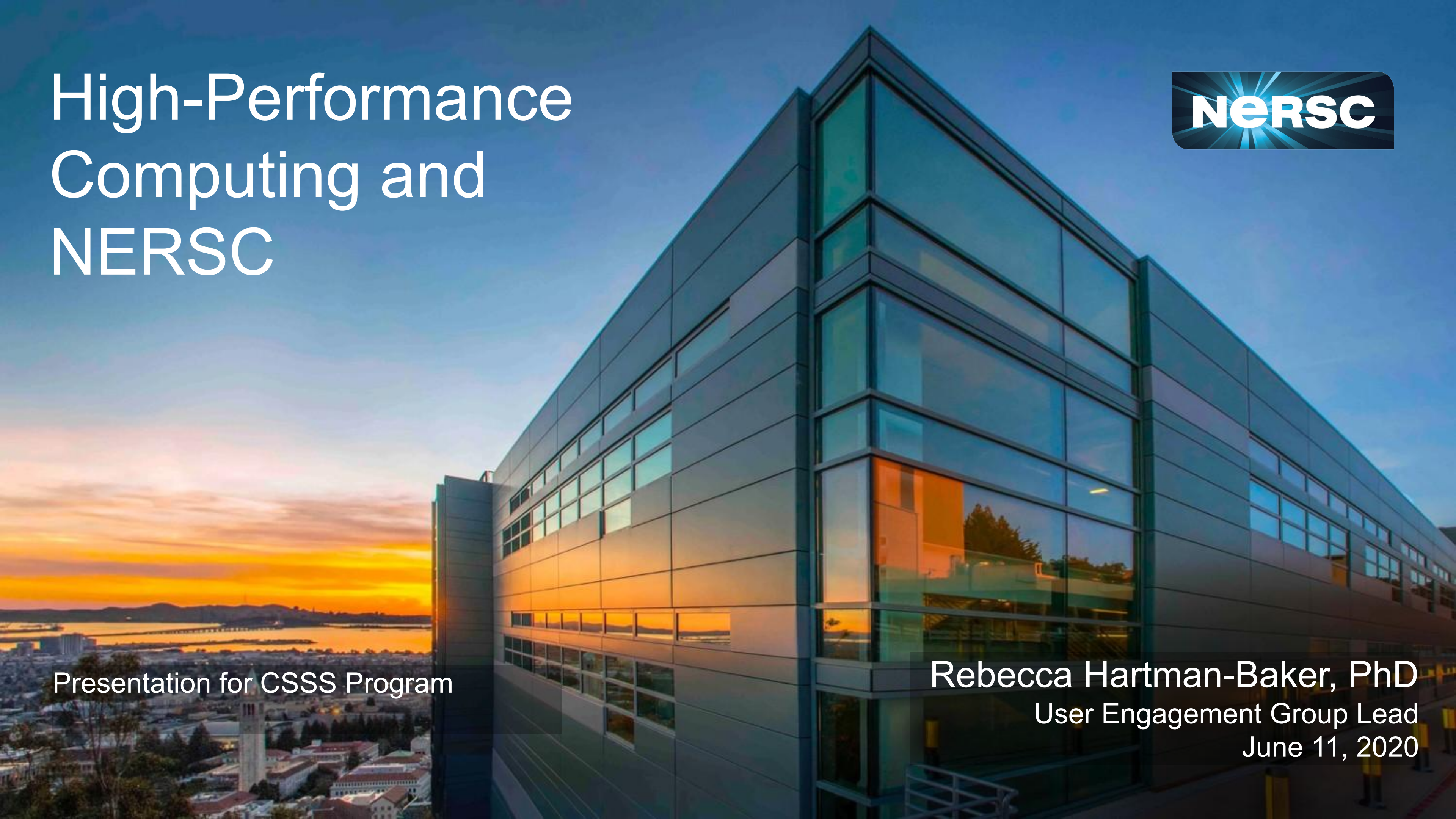


High-Performance Computing and NERSC

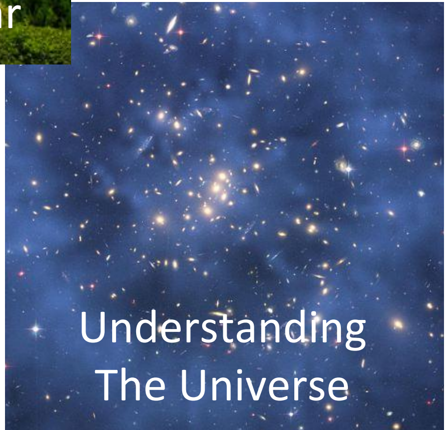
Presentation for CSSS Program

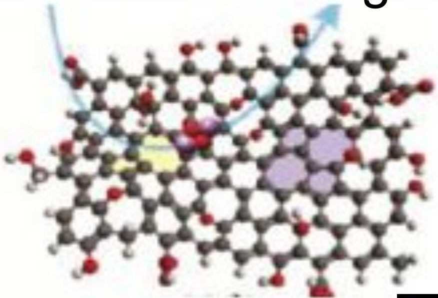
Rebecca Hartman-Baker, PhD
User Engagement Group Lead
June 11, 2020

High-Performance Computing Is...

… the application of "supercomputers" to scientific computational problems that are either too large for standard computers or would take them too long.

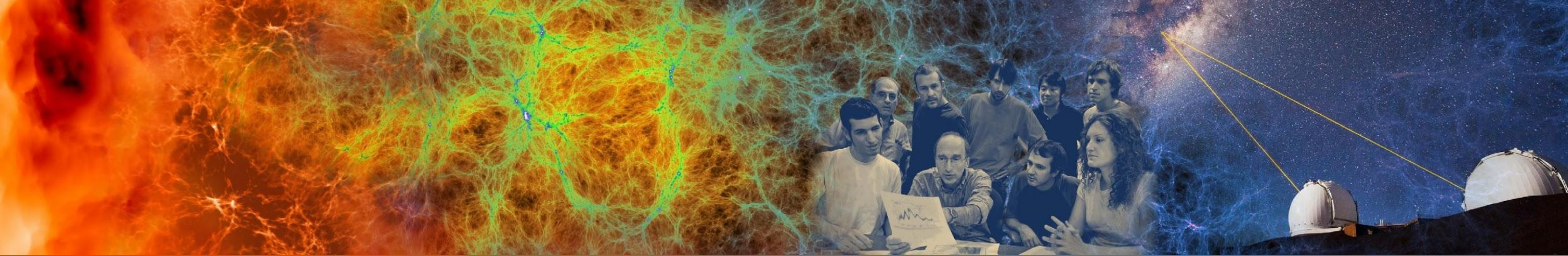

Cheap & Efficient Solar


Understanding The Universe


Designing Better Batteries


Understanding How Proteins Work


Extreme Climate Events



What Is a Supercomputer?

What Is a Supercomputer?

A. A processor (CPU) unimaginably more powerful than the one in my laptop.

B. A quantum computer that takes advantage of the fact that quantum particles can simultaneously exist in a vast number of states.

C. Processors not so different than the one in my laptop, but 100s of thousands of them working together to solve a problem.

A Supercomputer Is...

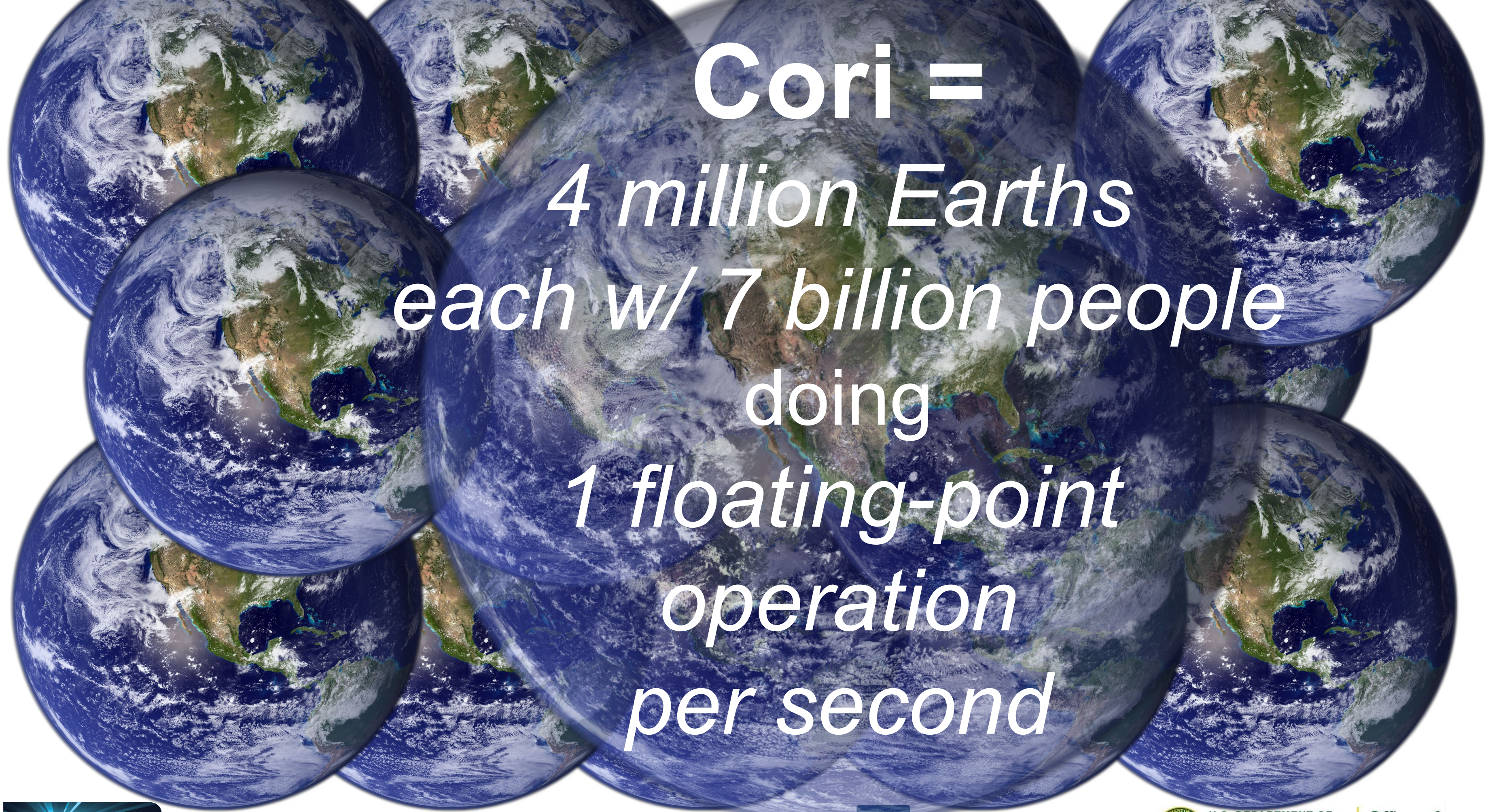
vs.

… not so different from a super high-end desktop computer.

Or rather, a lot of super high-end desktop computers.

Cori (left) has ~11,000 nodes (~ high-end desktop computers)

700,000 compute cores that can perform ~3x10^{16} calculations/second


Cori =
4 million Earths
each w/ 7 billion people
doing
1 floating-point
operation
per second

But Wait, There's More!

The nodes are all connected to each other with a high-speed, low-latency network.

This is what allows the nodes to "talk" to each other and work together to solve problems you could never solve on your laptop or even 150,000 laptops.

Typical point-to-point bandwidth

- Supercomputer: 10 GBytes/sec
- Your home: 0.02* GBytes/sec

5,000 X

Latency

- Supercomputer: 1 μs
- Your home computer: 20,000* μs

20,000 X

* If you're really lucky

Cloud systems have slower networks

...and Even More!

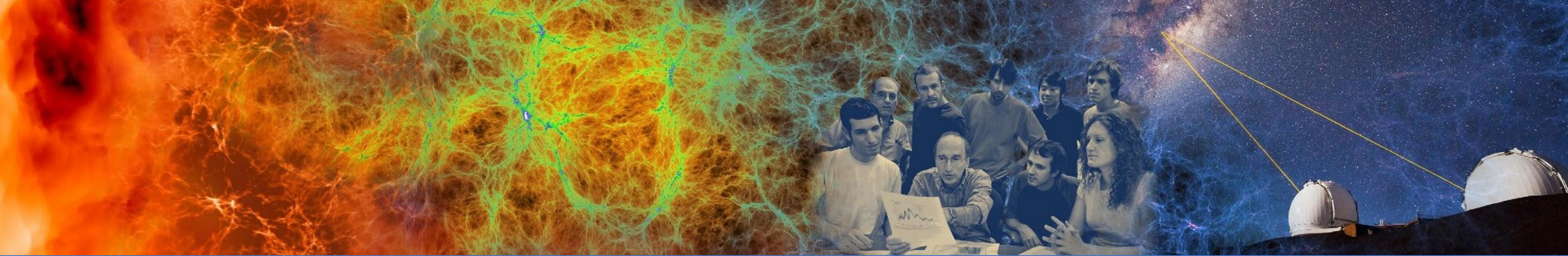
PBs of fast storage for files and data

- Cori: 30 PB
- Your laptop: 0.0005 PB
- Your iPhone: 0.00005 PB

Write data to permanent storage

- Cori: 700 GB/sec
- My iMac: 0.01 GB/sec

Cloud systems have slower I/O and less permanent storage



High-Performance Computing

High-Performance Computing...

- implies parallel computing
- In parallel computing, scientists divide a big task into smaller ones
- “Divide and conquer”

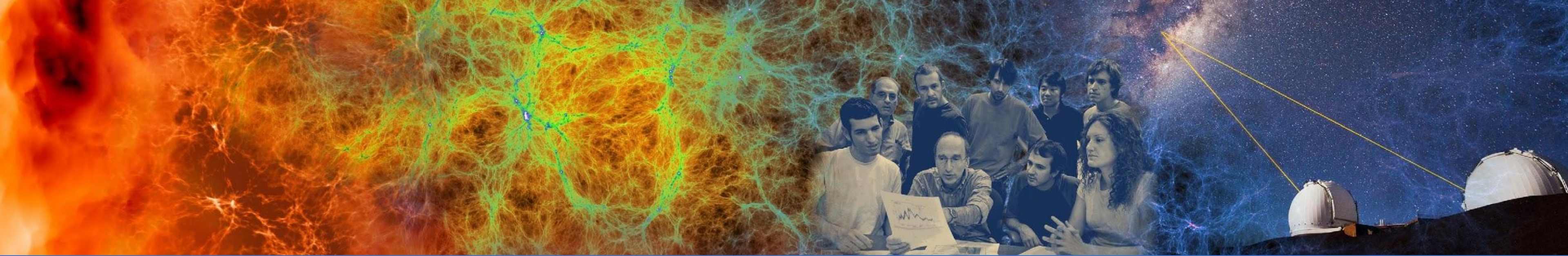
For example, to simulate the behavior of Earth’s atmosphere, you can divide it into zones and let each processor calculate what happens in each.

From time to time each processor has to send the results of its calculation to its neighbors.

Distributed-Memory Systems

This maps well to HPC “distributed memory” systems

- Many nodes, each with its own local memory and distinct memory space
- A node typically has multiple processors, each with multiple compute cores (Cori has 32 or 68 cores per node)
- Nodes communicate over a specialized high-speed, low-latency network
- SPMD (Single Program Multiple Data) is the most common model
    - Multiple copies of a single program (tasks) execute on different processors, but compute with different data
    - Explicit programming methods (MPI) are used to move data among different tasks



What is NERSC?

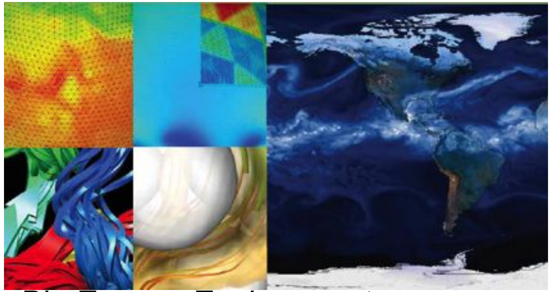
National Energy Research Scientific Computing Center

- NERSC is a national supercomputer center funded by the U.S. Department of Energy Office of Science (SC)
    - Supports SC research mission
    - Part of Berkeley Lab
- If you are a researcher with funding from SC, you can use NERSC
    - Other researchers can apply if research is in SC mission
- NERSC supports 7,000 users, 800 projects
    - From all 50 states + international; 65% from universities
    - Hundreds of users log on each day

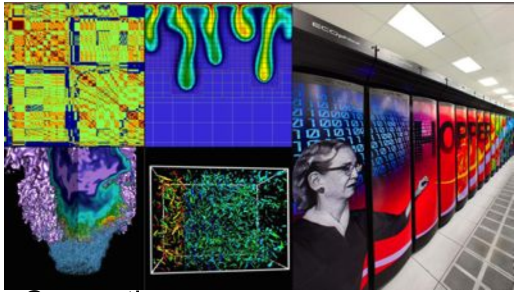
NERSC is the Production HPC & Data Facility for DOE Office of Science Research

U.S. DEPARTMENT OF ENERGY | Office of Science

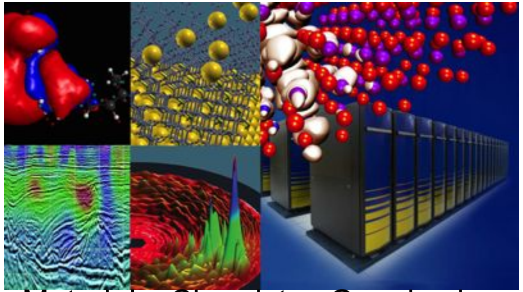
Largest funder of physical science research in U.S.



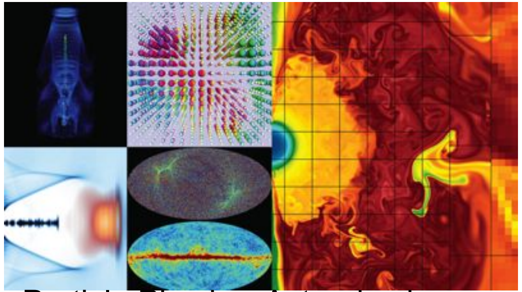
Bio Energy, Environment



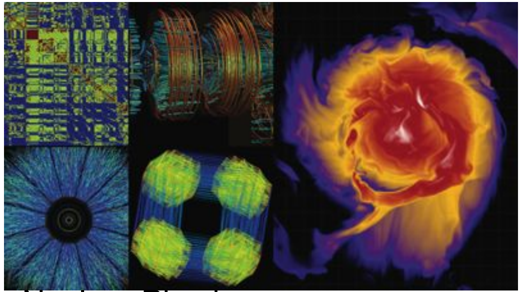
Computing



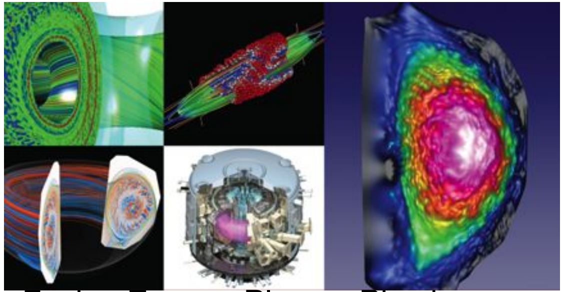
Materials, Chemistry, Geophysics



Particle Physics, Astrophysics



Nuclear Physics



Fusion Energy, Plasma Physics

NERSC: Science First!

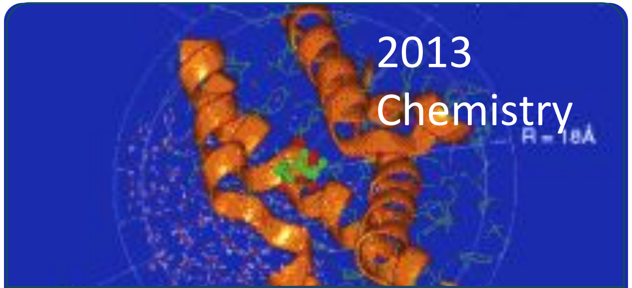
NERSC's mission is to accelerate
scientific discovery
at the DOE Office of Science
through high-performance computing
and data analysis.

2018 Science Output

>2500 refereed publications

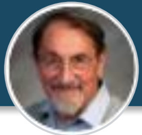
- Nature (14),
  Nature Communications (31),
  Other Nature journals (37)
- Science (11),
  Science Advances (9)
- Proceedings of the National Academy of Sciences (31)
- Physical Review Letters (67),
  Physical Review B (85)

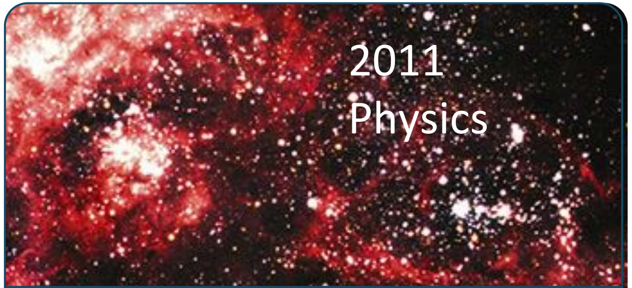
NERSC Nobel Prize Winners



2013
Chemistry

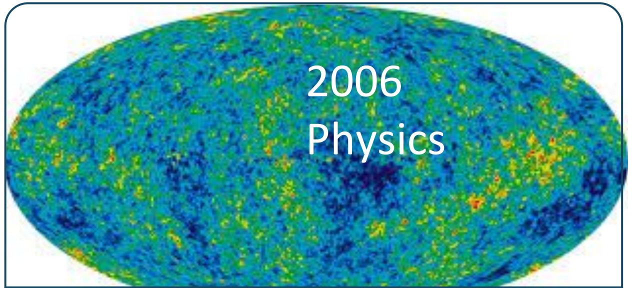
Martin Karplus





2011
Physics

Saul Perlmutter



2006
Physics

George Smoot

2007
Peace

Warren Washington

2015 Nobel Prize in Physics

Scientific Achievement

The discovery that neutrinos have mass & oscillate between different types

Significance and Impact

The discrepancy between predicted & observed solar neutrinos was a mystery for decades. This discovery overturned the Standard Model interpretation of neutrinos as massless particles and resolved the “solar neutrino problem”

Research Details

The Sudbury Neutrino Observatory (SNO) detected all three types (flavors) of neutrinos & showed that when all three were considered, the total flux was in line with predictions. This, together with results from the Super Kamiokande experiment, was proof that neutrinos were oscillating between flavors & therefore had mass.


A SNO construction photo shows the spherical vessel that would later be filled with water.

Calculations performed on PDSF & data stored on HPSS played a significant role in the SNO analysis. The SNO team presented an autographed copy of the seminal *Physical Review Letters* article to NERSC staff.

Q. R. Ahmad et al. (SNO Collaboration). Phys. Rev. Lett. 87, 071301 (2001)

Nobel Recipients: Arthur B. McDonald, Queen’s University (SNO)
Takaaki Kajita, Tokyo University (Super Kamiokande)

How California Wildfires Can Impact Water Availability

Scientific Achievement

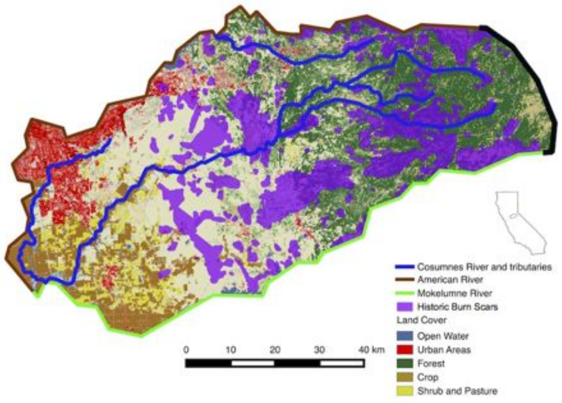
Berkeley Lab researchers used NERSC supercomputers to show that conditions left behind by California wildfires lead to greater winter snowpack, greater summer water runoff and increased groundwater storage.

Significance and Impact

In recent years, wildfires in the western United States have occurred with increasing frequency and scale. Even though California could be entering a period of prolonged droughts with potential for more wildfires, there is little known on how wildfires will impact water resources. The study is important for planners and those who manage California's water.

Research Details

The researchers modeled the Cosumnes River watershed, which extends from the Sierra Nevadas down to the Central Valley as a prototype of many California watersheds. Using about 3 million hours on NERSC's Cori supercomputer to simulate watershed dynamics over a period of one year the study allowed them to identify the regions that were most sensitive to wildfire conditions, as well as the hydrologic processes that are most affected.



Berkeley Lab researchers built a numerical model of the Cosumnes River watershed, extending from the Sierra Nevada mountains to the Central Valley, to study post-wildfire changes to the hydrologic cycle. (Credit: Berkeley Lab).

Maina, FZ, Siirila-Woodburn, ER. Watersheds dynamics following wildfires: Nonlinear feedbacks and implications on hydrologic responses. Hydrological Processes. 2019; 1– 18. https://doi.org/10.1002/hyp.13568

NERSC Project PI: Erica Siirila-Woodburn, Lawrence Berkeley National Laboratory

DOE Mission Science, Funded by UC National Laboratory Fees Research Program

Mapping Neutral Hydrogen in the Early Universe

Scientific Achievement

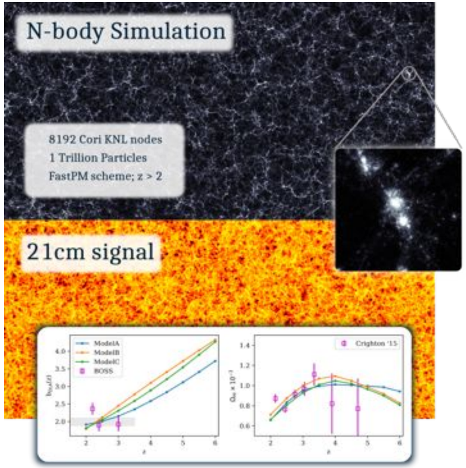
Researchers at the Berkeley Center for Cosmological Physics developed a model that produces maps of the 21 cm emission signal from neutral hydrogen in the early universe. Thanks to NERSC supercomputers, the team was able to run simulations with enough dynamic range and fidelity to theoretically explore this uncharted territory that contains 80% of the observable universe by volume and holds the potential to revolutionize cosmology.

Significance and Impact

One of the most tantalizing, and promising cosmic sources is the 21 cm line in the very early universe. This early time signal combines a large cosmological volume for precise statistical inference, with simple physics processes that can be more reliably modeled after the cosmic initial conditions. The model developed in this work is compatible with current observational constraints, and serves as a guideline for designing intensity mapping surveys and for developing and testing new theoretical ideas.

Research Details

The team developed a quasi-N-body scheme that produces high-fidelity realizations of dark matter distribution of the early universe, and then developed models that connects the dark matter distribution to the 21cm emission signal from neutral hydrogen. The simulation software FastPM was improved to run the HiddenValley simulation suite, which employs 1 trillion particles each, and runs on 8,192 Cori KNL nodes – **the largest N-body simulation ever carried out at NERSC**.



Upper panel: dark matter with an inset of the most massive galaxy system in the field of view. Lower panel: 21cm emission signal with an inset of the clustering properties compared with current constraints. Horizontal span: 1.4 comoving Gpc (6 billion light years); Thickness: 40 million light years.

Modi, Chirag; Castorina, Emanuele; Feng, Yu; White, Martin, "; Journal of Cosmology and Astroparticle Physics 2019 Sep, 10.1088/1475-7516/2019/09/024

NERSC Project PI: Yu Feng (UC Berkeley)
NERSC Director's Reserve Project, Funded by University of California, Berkeley

BERKELEY LAB
Bringing Science Solutions to the World

U.S. DEPARTMENT OF ENERGY

Machine-Learned Impurity Prediction in Semiconductors

Scientific Achievement

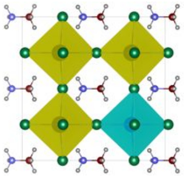
Argonne National Laboratory researchers ran high-throughput simulations on NERSC supercomputers and generated comprehensive datasets of impurity properties in two classes of semiconductors: lead-based hybrid perovskites and cadmium-based chalcogenides. These datasets led to machine learned models that enable accelerated prediction and design for the entire chemical space of materials and impurities in these semiconductor classes.

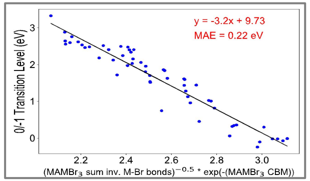
Significance and Impact

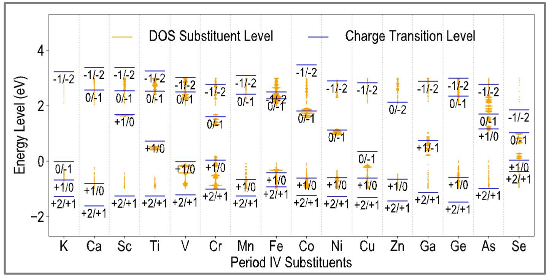
Impurity energy levels in semiconductors can change their behavior in ways that have important consequences for solar cell applications. The ability to instantly and accurately estimate such impurity levels is paramount. The current research combines simulation and machine learning to generate results that can potentially transform the design of novel semiconductors that are defect-tolerant or have tailored impurity properties.

Research Details

The researchers performed density functional theory calculations for hundreds of impurity atoms in selected semiconductors to determine their formation enthalpies and energy levels. The results were transformed into predictive models using machine learning algorithms. The DFT simulations modeled systems containing ~ 100 atoms, using ~ 1.5 million CPU hours.







High-throughput DFT data was generated for impurity energy levels in semiconductors (example shown for a hybrid perovskite above), which lead to machine-learned predictive models.

A. Mannodi-Kanakkithodi et al., "Comprehensive Computational Study of Partial Lead Substitution in Methylammonium Lead Bromide", accepted, Chem. Mater. doi: 10.1021/acs.chemmater.8b04017 (2019).

D.H. Cao et al., "Charge Transfer Dynamics of Phase Segregated Halide Perovskite mixtures", ACS Appl. Mater. Interfaces, 11 (9), pp 9583–9593 (2019).

NERSC Project PI: Maria K.Y. Chan, Argonne National Lab

DOE Mission Science, Funded by Basic Energy Sciences, Office of Energy Efficiency and Renewable Energy

NERSC

BERKELEY LAB
Bringing Science Solutions to the World

U.S. DEPARTMENT OF ENERGY

BERKELEY LAB

NERSC Usage Facts & Figures

In 2019, scientists used
8,770,000,000
NERSC-hours
and currently store
220,000,000
GB of data
at NERSC

>1,000,000 single-CPU-years

4 million iPhones

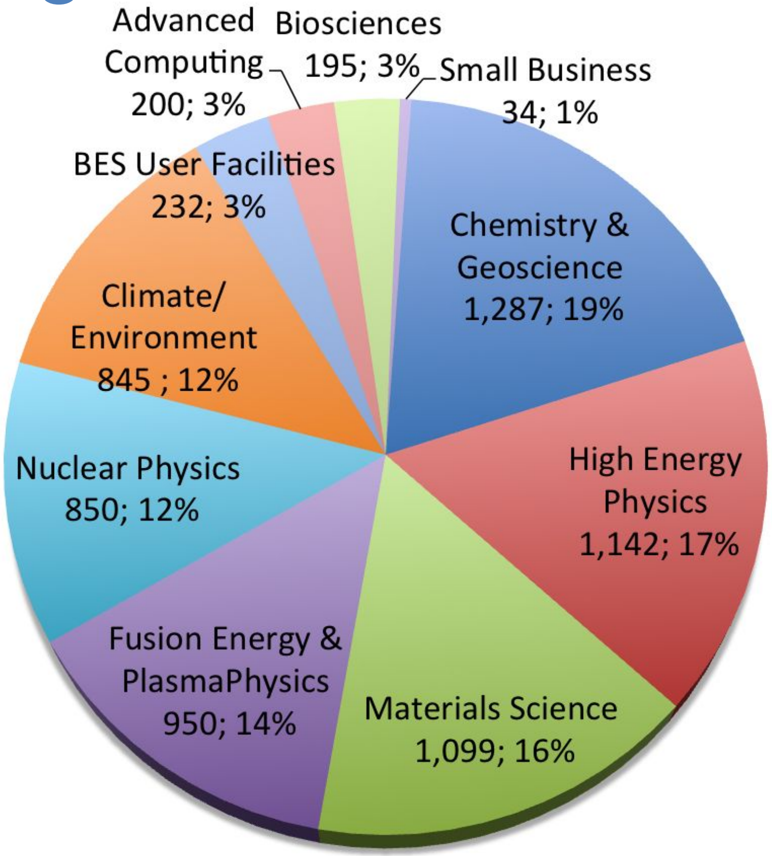
Homo erectus
~1,000,000 years ago

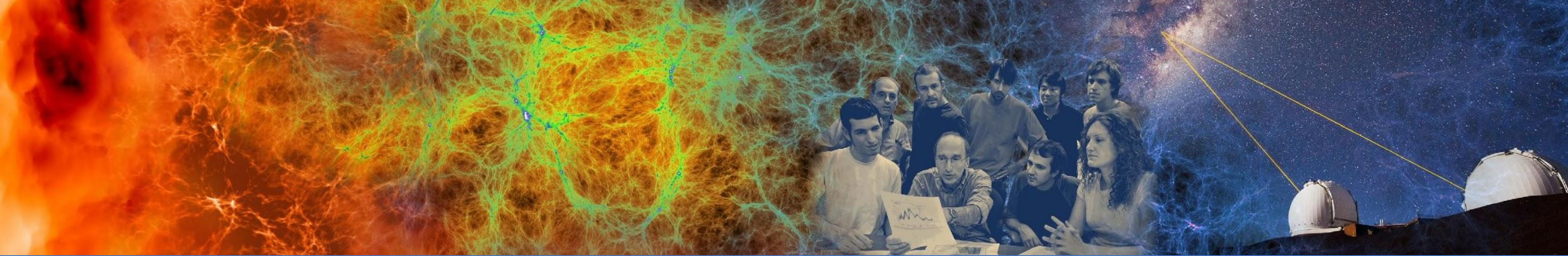
Data Storage

HPSS
200 Petabytes

Community
64 Petabytes

Compute Usage

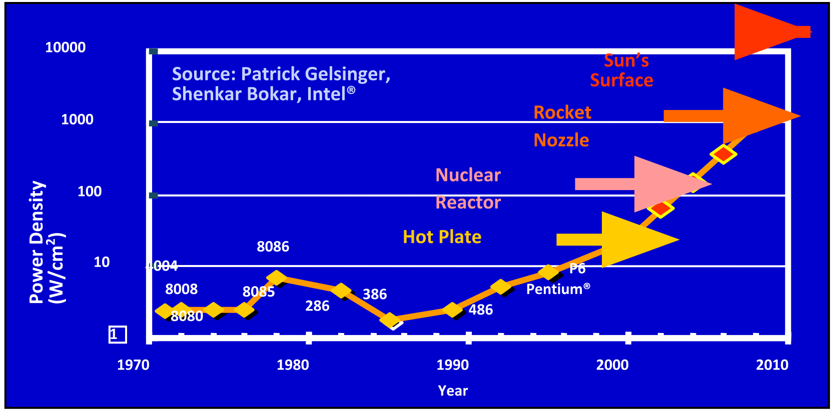


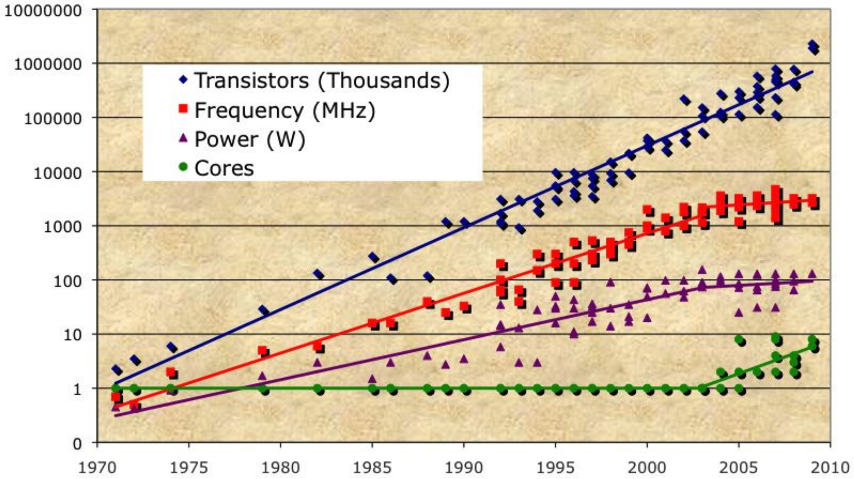


Challenges in HPC

Power: the Biggest Architectural Challenge

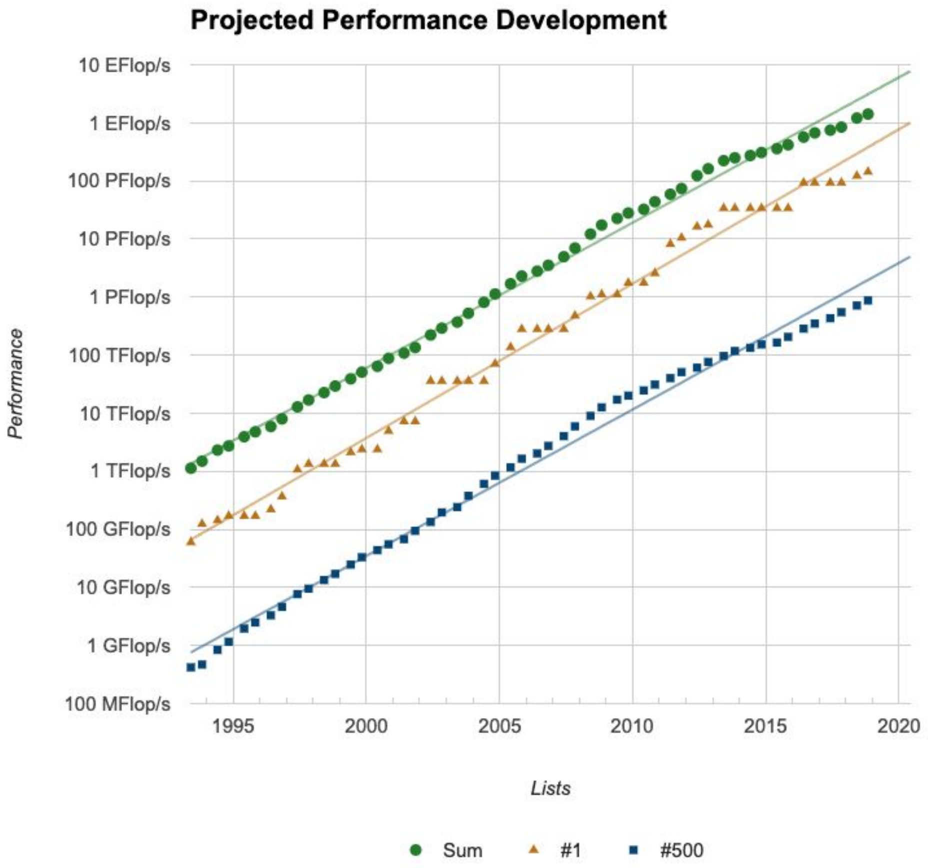
- If we just kept making computer chips faster and more dense, they'd melt and we couldn't afford or deliver the power.
- Now compute cores are getting slower and simpler, but we're getting lots more on a chip.
    - GPUs and Intel Xeon Phi have 60+ "light-weight cores"





Revolution in Energy Efficiency Needed

- Energy efficiency is increasing, but today's top computers use 10s of Megawatts of power at ~$1M/MW.
- Power bill for an Exascale machine made with today's tech exceeds budget for machine!



Programming for Advanced Architectures

- Advanced architectures (e.g., CPU+GPU offload) present challenges in programming and performance
  - Science expert must become expert on computer architectures and programming models
  - Performance on one architecture doesn't always translate to performance on another
  - Many codes not ported and many unsuitable for this type of architecture; complete overhaul required

Beyond Moore's Law

- Moore's law: doubling of performance every 18-24 months
  - There is an end, and it is soon
  - What do we do next?
- Pathfinding new architectures
  - Accelerators? FPGAs? Quantum?
  - How to program for these?

Data: Getting Bigger All the Time!

- Simulations producing more data
- Scientific instruments producing more data
  - SKA when comes fully online will produce more data in a day than currently exists!
- How do we
  - process this data?
  - manage it?
  - store it?
  - transfer it?
  - access it?
- Efficient workflows for data analysis and management needed

Your Challenges

- Figure out how to program the next generation of machines
- Find a way to make sense of all the data
- Build faster, more capable hardware that uses less energy
- Create effective data and job management workflows
- Bring new fields of science into HPC
- Tell the world about what you’re doing!

Questions?